

## シングルセル遺伝子発現データを利用した遺伝子ネットワークの構築手法 — 「疎」なデータから遺伝子群に内包される正負の相関を推定する —

### 1. 発表者：

仲嶋 なつ（東京大学定量生命科学研究所 大規模生命情報解析研究分野 特任研究員（研究当時）／東京大学医科学研究所 特任研究員（現在））  
林 寛敦（東京大学定量生命科学研究所 分子病態情報学社会連携部門 特任助教）  
藤木 克則（東京大学定量生命科学研究所 ゲノム情報解析研究分野 助教）  
白髭 克彦（東京大学定量生命科学研究所 ゲノム情報解析研究分野 教授）  
秋山 徹（東京大学定量生命科学研究所 分子病態情報学社会連携部門 特任教授）  
阿久津 達也（京都大学化学研究所 バイオインフォマティクスセンター 教授）  
中戸 隆一郎（東京大学定量生命科学研究所 大規模生命情報解析研究分野 講師）

### 2. 発表のポイント：

- ◆ 「疎」でありデータのばらつきが多いシングルセル遺伝子発現データから、頑健に遺伝子ネットワーク推定や排他的発現パターンを示す遺伝子ペアを同定する新規手法を開発しました。
- ◆ 構築した手法を膠芽腫幹細胞に対して適用し、従来法では見つけられなかった新規の腫瘍マーカー遺伝子候補を複数同定しました。
- ◆ 本手法はどのような生体組織データに対しても利用可能であり、新規の重要なマーカー遺伝子の発見に活用されることが期待されます。

### 3. 発表概要：

シングルセル遺伝子発現量解析（scRNA-seq）は、生体組織や腫瘍組織に含まれる細胞不均一性（注1）を同定するための強力な手法です。一方でデータ感度は限られており、ばらつきが多い「疎」なデータであるため、従来の遺伝子間相関・ネットワーク推定手法をそのまま適用することが困難でした。東京大学定量生命科学研究所の仲嶋 なつ 特任研究員、中戸 隆一郎 講師らの研究グループは、そのようなscRNA-seqデータに対して頑健に遺伝子共発現ネットワーク（注2）を構築・比較する新規手法を開発しました。さらに、相互排他的な遺伝子（注3）の推定を行うための指標、Exclusively Expressed Index（EEI）を提案し、相互排他的な遺伝子ペア群を細胞クラスタリングのための特徴量として応用する手法を開発しました。

本手法を同研究所の林 寛敦 特任助教、秋山 徹 特任教授らの研究グループによって樹立された膠芽腫幹細胞データに適用し、血清刺激前後において共発現・排他発現パターンが変動する遺伝子群を網羅的に同定しました。これにより、従来法では見つけられなかった新規の膠芽腫幹細胞マーカー遺伝子候補を複数同定しました。本手法を活用することで、種々の組織細胞における新規マーカー遺伝子の同定や、細胞不均一性への理解の向上につながることが期待されます。

本研究の成果は、日本学術振興会 科研費 新学術領域研究「細胞社会ダイバーシティの統合的解明と制御」（計画班代表者：中戸隆一郎、秋山徹、分担者：林寛敦）などの支援によって得られたものです。

#### 4. 発表内容：

生体組織や腫瘍組織などに含まれる細胞群は均一ではなく、様々な種類の細胞によって多様な「細胞ダイバーシティ」を形成しています。この細胞ダイバーシティの形成・変化が、腫瘍の薬剤耐性などにも関係していると考えられています。近年の技術的進歩により、生体組織のゲノム情報を個々の細胞レベルで網羅的に計測する「シングルセル解析」が可能となりました。シングルセル解析では細胞群の平均的な挙動だけではなく、細胞群が持つ細胞不均一性や、内包されるそれぞれの細胞種が持つ性質（例えば薬剤応答性など）を詳細に個別に分析することができます。シングルセル遺伝子発現量解析（scRNA-seq）は一細胞レベルで遺伝子発現量を計測する手法であり、世界中で最もよく用いられているDrop-seq法を用いた方法では、数千から数万細胞もの遺伝子発現量を一度に明らかにすることが可能です。一方で、技術的な制約から一細胞あたりのデータ量は十分ではなく、得られたデータのばらつきが大きくなり、また発現しているはずの遺伝子の発現量が0と計測されてしまう、いわゆる「疎」なデータとなり、従来用いられる情報解析手法である遺伝子間相互作用推定などの手法が適用しにくいという課題がありました。

今回、研究グループは、細胞ダイバーシティにおいて重要な役割を担う未知の遺伝子を探索するための新規手法を開発しました。疎なデータに対して頑健に遺伝子間の相関を推定するCo-Dependency Index（CDI）という指標を導入し、scRNA-seqデータから遺伝子共発現ネットワーク（注2）を構築する手法を開発しました。CDIは任意の遺伝子ペアが共発現している可能性を確率的に計算する手法であり、低発現遺伝子の発現パターンを効率的に捉えることができます。また、そうして得られたネットワークからGirvan-Newman法、Leading eigenvector法を用いたモジュール抽出（注4）を行い、細胞ダイバーシティにおける共発現遺伝子群を推定する手法を提案しました。本手法を血清刺激前後の膠芽腫幹細胞データに適用し、血清刺激前と刺激後のサンプル間でモジュール比較を行った結果、既知の幹細胞マーカーであるPDGFRAとMETが異なるモジュールに含まれており、それぞれ異なる共発現遺伝子群を持つこと、血清刺激によってPDGFRAの発現は失われるものの、その共発現遺伝子の一部分は発現が失われないことを発見しました。このことは、モジュール単位でのサンプル間比較が、共発現遺伝子群における類似性や相違性を明らかにすること、個々の遺伝子に着目した発現解析では得られないマーカー遺伝子の同定に有効であることを示唆しています。

更に研究グループは、相互排他的な遺伝子ペアの推定を行うための新規指標であるExclusively Expressed Index（EEI）を開発しました。細胞ダイバーシティにおいては共発現と同様に、排他的遺伝子発現も本質的に重要です。特に一細胞レベルでは、遺伝子変異による排他的発現が細胞不均一性を引き起こす要因の一つではないかと考えられます。EEIはCDIの原理を応用し、2種類の排他的な状態を考慮した指標であり、細胞ダイバーシティにおいて排他的な発現パターンを示す遺伝子ペアを網羅的に同定可能です。さまざまなscRNA-seqデータを用いて精度評価実験を行った結果、EEIは他手法と比べて一細胞あたりのデータ量がより減少した場合でも相互排他的な遺伝子ペアを頑健な推定することが可能であり、マーカー遺伝子特定の有効性を明らかにしました。ここでは更に、EEIを用いて推定された相互排他的な遺伝子ペア群を、細胞クラスタリングのための特徴量として応用する手法を開発しました。本手法をプロトコルの異なるscRNA-seqデータに適用した結果、EEIで推定された排他的な遺伝子ペアを特徴量として追加することで、クラスタリング精度が向上することが確認されました。この結果は、排他的遺伝子ペア群を特徴量として応用することで、検出感度や細胞不均一性への理解の向上につながることを、提案手法が実験プロトコルやデータ量の違いによらず適用できる頑健性を持つことを示唆しています。

本研究は、世界中で最もよく用いられているscRNA-seqデータに対しても利用可能な、遺伝子群の共発現性と排他性を頑健に推定する手法を提案しました。本手法を活用することで、種々の組織細胞における新規マーカー遺伝子の同定や、細胞不均一性への理解の向上につながることを期待されます。

## 5. 発表雑誌：

雑誌名：Nucleic Acids Research

論文タイトル：Codependency and mutual exclusivity for gene community detection from sparse single-cell transcriptome data

著者：Natsu Nakajima, Tomoatsu Hayashi, Katsunori Fujiki, Katsuhiko Shirahige, Tetsu Akiyama, Tatsuya Akutsu and Ryuichiro Nakato\*

DOI 番号：10.1093/nar/gkab601

URL：https://doi.org/10.1093/nar/gkab601

## 6. 問い合わせ先：

東京大学定量生命科学研究所 大規模生命情報解析研究分野

講師 中戸 隆一郎（なかと りゅういちろう）

TEL: 03-5841-1471 FAX: 03-5841-7308

E-mail: rnakato@iqb.u-tokyo.ac.jp

## 7. 用語解説：

（注1）細胞不均一性：

生体組織や腫瘍組織では、性質の異なる多様な細胞種から構成されている。シングルセル解析では遺伝子発現量パターンなどをもとに細胞種の分類（クラスタリング）が行われる。

（注2）遺伝子ネットワーク：

遺伝子間の相互作用を、遺伝子発現データと数理的手法に基づいて解析し、表現したネットワーク図。遺伝子を表す頂点と、相互作用を表す辺により構成される。ここでは全細胞を通じて発現パターンが類似している（共発現している）遺伝子間に辺を作成している（遺伝子共発現ネットワーク）。

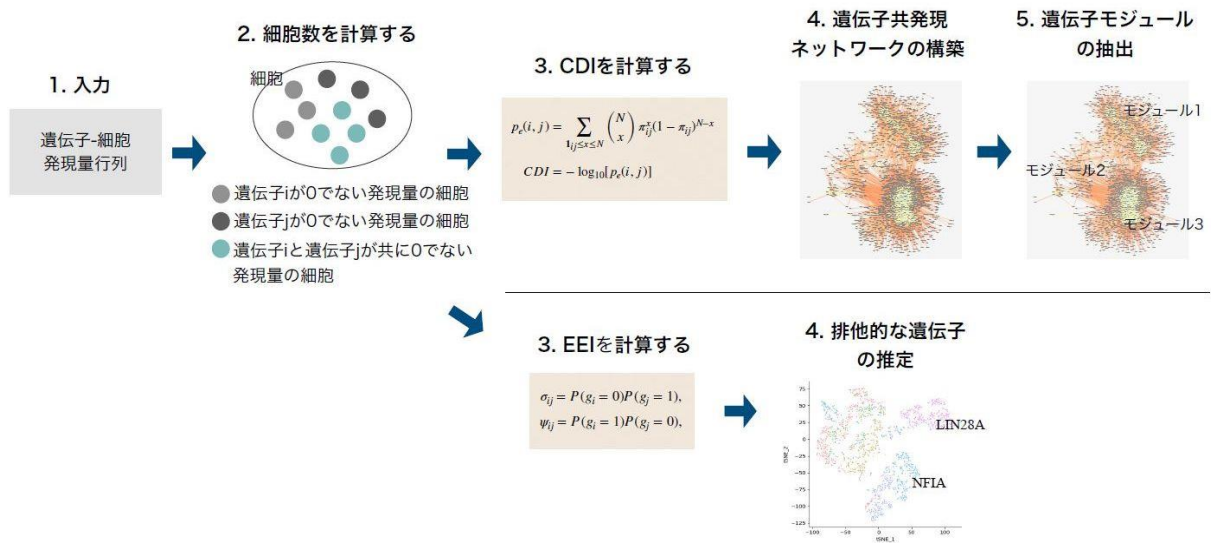
（注3）相互排他的な遺伝子：

ある細胞では高発現を示し、ある細胞では低発現を示す遺伝子ペア、もしくは、逆相関性を持つ遺伝子ペア。

（注4）モジュール：

遺伝子ネットワークから抽出される部分的なネットワーク。

## 8. 添付資料：



図：提案手法では、シングルセル遺伝子発現量データ (scRNA-seq) を入力として、共発現、排他的発現パターンを示す遺伝子ペアを網羅的に計算します。CDIを用いて、遺伝子共発現ネットワークの構築と遺伝子モジュールの抽出を行います。EEIを用いて排他的な遺伝子ペアを推定します。この排他的遺伝子ペアは細胞クラスタリングの特徴量として利用可能です。